

CASE STUDY – Translated from an article in 01 Informatique magazine

Cartographic Conservation at the IGN

In order to create a hierarchical and secure archive for their dozens of terabytes of image and cartographic data, the IGN has chosen a solution from the software vendor Active Circle.

© IGN

ORGANIZATION

Institut Géographique National (IGN)

ACTIVITY: production and sale of maps and geographical databases

HQ: Vincennes, near Paris, France

PERSONNEL: approx. 1,700

PROBLEM TO RESOLVE

Archive all of the geographical data for the IGN portal.

SOLUTION DEPLOYED

Active Circle, a storage and archiving solution. Overland NEO 8000 Tape Library.

THE IGN AND THE SIEL

The On-Line Information Service (SIEL) is responsible for archiving all the cartographic data produced by the IGN. Additionally, the Service is also in charge of managing the images available via their website Geoportail (<http://www.geoportail.fr>). This involves the management of two different types of data: vector images for the maps and raster images stored in TIFF and ECW (Enhanced Compression Wavelet) formats. "The production teams at the IGN send us their data in SIP (Submission Information Package) format, which includes images that have already been compressed in TAR format and the associated metadata. Obviously, the data streams are much larger for the raster images," notes Yann Le Disez, manager of the distribution and archiving department at the SIEL.

Previously this IGN department employed the Dorostore system in conjunction with an electronic catalog supported externally, resulting in very long archiving times. In 2007, in order to confront the extraordinary growth of data to archive (50 Tb through 2007 with an additional 25 Tb estimated at the time for 2008), the SIEL decided to completely re-engineer the system, opting for a unified, secure, and scalable archiving solution on LTO tape. An invitation to tender was announced and four system integrators responded.

REQUIREMENTS: ADAPTING TO BEST PRACTICES

In the end, it was the solution proposed by the system integrator Komposite that won out over the three others. For the SIEL, it was the one best adapted to the specific developments already implemented. "In conformity with the procedures of a European open invitation to tender, we had defined three criteria for selection: technology, price, and delivery schedule. One very well-known integrator proposed a system with all the required elements, but which we had to reject because we wanted an open tool capable of integrating with the elements of our existing workflow," explained Yann Le Disez.

The department archives its data according to well-defined procedures using the Open Archival Information System (OAIS) conceptual model. This process requires that before creating any archive they must generate the previously mentioned SIPs, which are based on an open compression standard, rather than relying on the compression functionality and capabilities of a proprietary archiving system. This operation is in fact carried out in the individual production departments using a program developed in-house.

ARCHITECTURE: A FAULT-TOLERANT SYSTEM

Two high-performance NAS units from Exanet host the cartographic data that can be accessed from the site according to well-defined hierarchical rules (retention times differ according to the content). They also serve as a buffer for the final archiving to LTO4 cartridges in the NEO 8000 tape library located at the IGN's Saint-Mandé site outside of Paris.





3D view of Marseille, France
© IGN - Virtual City

The Active Circle software manages the transfer of this data across its system of storage nodes. A node consists of the storage space of one of the NAS units (4 Tb of cache and 25 Tb of operational storage) and a partition of the NEO 8000 tape library. To guarantee the high level of fault tolerance, the SIEL has installed two nodes at its main site and a third synchronized node with a second tape library at the site in Villefranche-sur-Cher, 200 kilometers to the south. As a result – and it is here that we see one of the advantages of the Active Circle cluster-based system – if one of the nodes goes down, the system continues to function on the two other machines while the problem is resolved.

OPERATION: VALIDATION REQUIRED

In practice, the procedure is more complicated than that. The ready-to-archive files submitted by the production teams are copied from the production servers and deposited into shared folders on the NAS unit. These folders are regularly scanned for changes by a software agent that launches a script to verify the SIPs (integrity and metadata validation). If the package is incorrect, an error message is sent by mail and the file is quarantined: it does not move to the next step and must be regenerated by the production team. If it is correct, Active Circle takes over and the file is “pushed” to one of the system’s main cells.

OBSTACLES: RESTRICTIVE INTERNAL PROCESSES

Although the Active Circle solution met nearly all of the SIEL’s requirements, there were a few points left uncovered. “The capability to validate the MD5 checksum that accompanies the SIP was not integrated in Active Circle,” notes Yann Le Disez. This was finally achieved with the support of the system integrator. Another obstacle, though not one presented by Active Circle, was the capacity of the tape cartridges: the system could not archive files larger than 800 Gb in size. To overcome this, it was decided to split large files in two and to archive them to two separate tapes. This is not managed by Active Circle but is accomplished upstream by the production teams.

PROJECT CALENDAR

Aug. 2007	End of requirements analysis and publication of invitation to tender.
Dec. 2007	Solution chosen, internal data center ready.
Feb. 2008	Deployment of Active Circle V.3
End 2009	35 Tb added to 25 Tb generated in 2008 and to 50 Tb from original system.
2010	50 Tb expected to be archived
2011	100 Tb expected to be archived.



Yann Le Disez, manager of the distribution and archiving department at the SIEL, a division of the IGN.

“The system is well adapted to our growing requirements.”

“Since the implementation of Active Circle, we have migrated all the old archives, close to 50 Tb. With an average of about 1 Tb of data flowing in per week, we now have about 110 Tb of archived data. We occasionally need to restore files, about two or three times per month. The preparation of the data and administration are what require the most attention. As for the rest, once the work upstream has been accomplished, we only need to push the data and then let the archiving processes do their work.”

Created in 2002, Active Circle develops software for organizations that manage large volumes of data: video content, images, scientific or technical data, or user information. The Active Circle solution optimizes data lifecycle management while at the same time simplifying storage administration and reducing total cost of ownership.

“Active Circle” is a registered trademark of Active Circle S.A. Any other names or brands are mentioned solely for the purposes of identification and are the property of their respective owners.

© ACTIVE CIRCLE 2011 – This document may not be copied or reproduced without written permission.

www.active-circle.com

